

A Maximum Entropy Approach to Species Distribution Modeling

Miroslav Dudik

Department of Computer Science
Princeton University

Species distribution modeling is an important problem in ecology and conservation biology. In this problem, we are given a list of locations where a species was observed and a set of environmental variables for the region interest, e.g. elevation, soil type and annual rainfall. Based on these, we would like to predict which conditions are favored by the species. Two main challenges for machine learning are the small number of occurrence localities and a lack of information where the species was NOT found (negative examples). To address both of these in a statistically sound manner, we propose to use the maximum entropy approach.

In this talk, we describe the maximum entropy principle (maxent) and how it relates to the maximum likelihood. We describe a relaxation of maxent that gives rise to non-asymptotic performance guarantees which depend very moderately on the number or complexity of environmental variables. We also present a novel coordinate descent algorithm for computing maxent models. Finally, we mention how maxent can be used to address the problem of sample selection bias.